AD-A041 629   CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER   F/G 12/2
               AVERAGE DELAY IN QUEUES WITH NONSTATIONARY POISSON ARRIVALS,(U)
               MAY 77   S M ROSS   N00014-77-C-0299
UNCLASSIFIED   ORC-77-13   NL
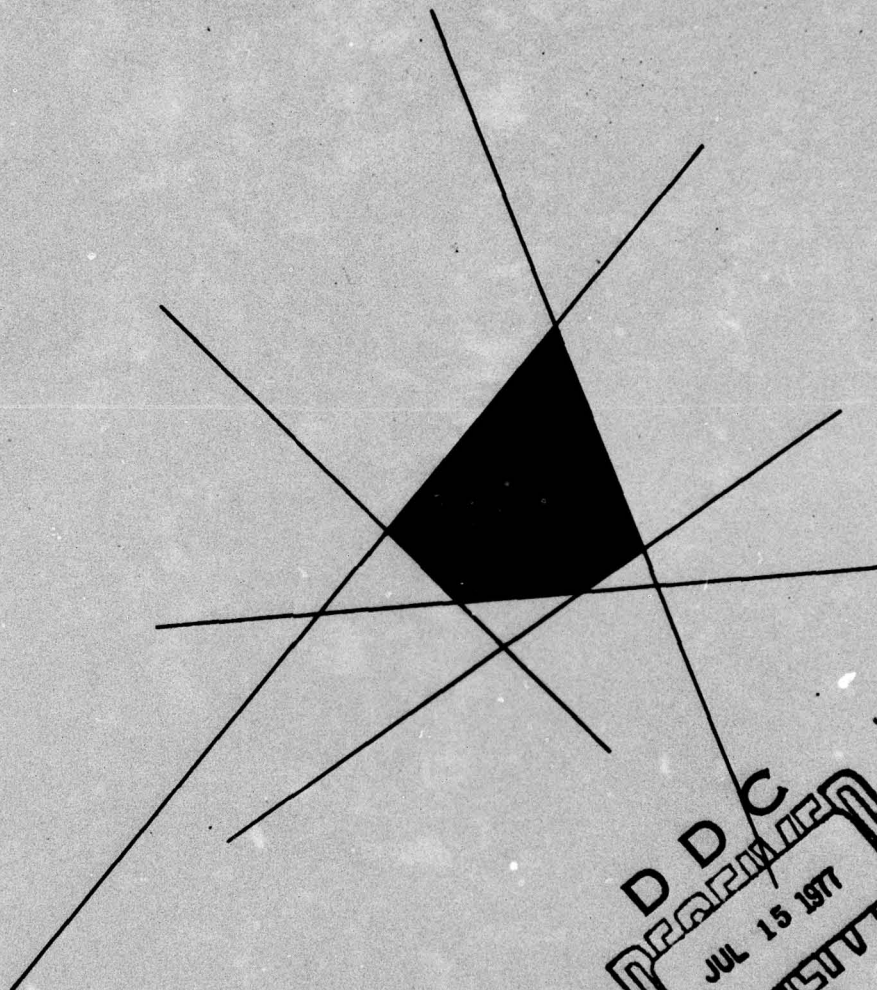
1 OF 1
ADA041629

END
DATE
FILMED
8-77

AVERAGE DELAY IN QUEUES WITH NONSTATIONARY POISSON ARRIVALS

by

SHELDON M. ROSS

C

$B.S.$

D D C

JUL 15 1977

C

OPERATIONS RESEARCH CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

AVERAGE DELAY IN QUEUES WITH NONSTATIONARY POISSON ARRIVALS

by

Sheldon M. Ross
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ORC-77-13	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AVERAGE DELAY IN QUEUES WITH NONSTATIONARY POISSON ARRIVALS.		5. TYPE OF REPORT & PERIOD COVERED Research Report,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Sheldon M. Ross		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0299
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California 94720		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042 379
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, Virginia 22217		12. REPORT DATE May 1977
		13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also supported by the Air Force Office of Scientific Research, AFSC, USAF,
under Grant AFOSR-77-3213.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonstationary Arrivals   Comparison with Stationary Case
Average Delay
General Service
Single Server
Monotonicity

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE

270754

ABSTRACT

One of the major difficulties in attempting to apply known
queueing theory results to real problems is that almost always
these results assume a time stationary Poisson arrival process,
whereas in practice the actual process is almost invariably
nonstationary. In this paper we considers single server in-
finite capacity queueing models in which the arrival process
is a nonstationary process with an intensity function $\Lambda(t)$,
$t \geq 0$, which is itself a random process. We suppose that the
average value of the intensity function exists and is equal to
some constant, call it λ , with probability 1.

We make a conjecture to the effect that the "closer $[\Lambda(t), t \geq 0]$
is to the stationary Poisson process with rate λ" then the smaller
is the average customer delay, and then we verify the conjecture is verified
in a special case.

iii

AVERAGE DELAY IN QUEUES WITH NONSTATIONARY POISSON ARRIVALS

by

Sheldon M. Ross

1. THE CONJECTURE

Consider a single server, infinite capacity, queueing model in which the arrival process is a nonstationary Poisson process with intensity function $\Lambda(t)$, $t \geq 0$; where $[\Lambda(t), t \geq 0]$ is itself a stochastic process. We suppose that $[\Lambda(t), t \geq 0]$ is such that

$$\lim_{t \to \infty} \int_0^t \frac{\Lambda(s)ds}{t} = \lambda \quad \text{(say)}$$

exists and is constant with probability 1. The above would follow, for instance, if $[\Lambda(t), t \geq 0]$ were a regenerative process with finite mean regeneration time T. In this case λ would be given by (see [2])

$$\lambda = E\left[\int_0^T \Lambda(s)ds\right] \bigg/ E[T] .$$

In addition we suppose that the successive service times are independently chosen from some service distribution G satisfying

$$E[S^2] \equiv \int x^2 dG(x) < \infty$$

and

$$E[S] \equiv \int x dG(x) < 1/\lambda .$$

The order of service is "first come first served."

Let us denote by d the average amount of time that a customer spends waiting in queue. We then make the following conjecture concerning d .

Conjecture:

$$d \geq \frac{\lambda E[S^2]}{2(1 - \lambda E[S])} \; .$$

It should be noted that $\frac{\lambda E[S^2]}{2(1 - \lambda E[S])}$ would be the value of d if the arrival process was a stationary Poisson process with rate λ .

In fact not only do we conjecture the above, but we also believe that the "closer $\{\Lambda(t), t \geq 0\}$ is to the stationary process with rate λ " then the smaller d is. For example suppose that $\{\Lambda(t), t \geq 0\}$ is a 2-state continuous time Markov chain alternating between the states λ_1 and λ_2 and suppose that the time it spends in state λ_i during each visit is an exponential random variable with rate $c\alpha_i$, $i = 1,2$. Let $d(c)$ denote the average customer delay for this model. It should be noted that the average arrival rate λ , given by

$$\lambda = \frac{\lambda_1 \alpha_2 + \lambda_2 \alpha_1}{\alpha_1 + \alpha_2} \; ,$$

is independent of the constant c which regulates how fast the arrival rate changes between its possible values. Intuitively the larger c is the "more stationary" the arrival process is and thus we conjecture that

$$d(c) \text{ is a decreasing function of } c$$

and

$$\lim_{c \to \infty} d(c) = \frac{\lambda E[S^2]}{2(1 - \lambda E[S])} \; .$$

We have only been able to prove the above conjecture in the special case $\lambda_2 = 0$. Before presenting this result, however, we need the following preliminary results on GI/G/1 queues.

2. A PRELIMINARY RESULT AND A COUNTEREXAMPLE

For any 2 probability distributions F and G we say that $F \leq_v G$ if

$$\int f(x) dF(x) \leq \int f(x) dG(x)$$

for all increasing convex functions.

Some easily derived properties of this ordering are:

1. $F \leq_v G$ if and only if

$$\int_a^\infty (1 - G(x)) dx \geq \int_a^\infty (1 - F(y)) dy \quad \text{for all } a \ .$$

2. If $F_i \leq_v G_i$ for $i = 1, 2$, then $F_1 * F_2 \leq_v G_1 * G_2$, where $*$ denotes convolution.

3. If $\int x dF(x) = \int x dG(x)$ then $F \leq_v G$ if

$$\int f(x) dF(x) \leq \int f(x) dG(x) \quad \text{for all convex functions } f \ .$$

If X and Y are random variables then we say that $X \leq_v Y$ if their distributions satisfy this ordering.

Remark:

If $E[X] = E[Y]$ then $X \leq_v Y$ intuitively means that X has less variability than Y .

We shall use the usual notation F/G/1 to describe the queueing system in which the arrival process is a renewal process with interarrival distribution F and there is a simple server having service distribution G. The following result was proven by Stoyan [2].

Proposition 1:

Consider the 2 queueing systems F/G/1 and $F^1/G^1/1$. If $F \leq_v F^1$ and $G \leq_v G^1$ and $\int x dF(x) = \int x dF^1(x)$ then $D_\infty \leq_v D_\infty^1$, where $D_\infty \left(D_\infty^1 \right)$ is the limiting distribution of delay in the system F/G/1 $(F^1/G^1/1)$.

The above can be proven by showing that $D_n \leq_v D_n^1$ where D_n and D_n^1 represent the delays in queue of the nth customers of the respective models. This is shown by induction on n, by using properties 2 and 3 (which implies that if $X \leq_v Y$ and $E[X] = E[Y]$ then $-X \leq -Y$) and the identity

$$D_{n+1} = \max [0, D_n + U_n] ,$$

where

$$U_n = \text{nth service time} - \text{nth interarrival time.}$$

Remark:

It is perhaps surprising that the analogue of Proposition 1 does not hold when there is more than one server. To best understand the forthcoming counterexample first consider 2 two-server queueing systems, the servers of the first system having constant service time of 1 unit, and the servers of the second system having service times that are either 0 or 2 - each possibility having probability 1/2. Suppose that in both systems customers

arrive in batches of size 3, with the times between each batch being large. Then it is easy to see that the average delay in the first (constant) system is 1/3 whereas it is 1/6 in the second (more variable) system.

The above example yields a counterexample even when we require the arrival process to be a renewal process. To see this suppose that the arrival process is a renewal process with interarrival times X_i where

$$X_i = \begin{cases} 0 & \text{with probability } \epsilon \\ \text{large} & \text{with probability } 1 - \epsilon . \end{cases}$$

Hence in this renewal arrival process the customers will arrive in batches of random size. As batches of size greater than 3 occur much more infrequently than those of size 3 [the probabilities being ϵ^3 versus $\epsilon^2(1 - \epsilon)$], it is clear from the preceding argument that average customer delay is greater in the constant service model.

3. PROOF OF THE CONJECTURE WHEN $\lambda_2 = 0$

Consider the infinite capacity single server model of Section 1 where $\Lambda(t)$ alternates between λ_1 and λ_2 , spending an exponential amount of time with rate $c\alpha_i$ during each visit to λ_i , $i = 1,2$. Let $D(c)$ denote the limiting distribution of customer delay in this system, and let $d(c)$ denote the mean of this distribution.

Proposition 2:

If $\lambda_2 = 0$ then $D(c_1) \underset{v}{\leq} D(c_2)$ whenever $c_1 \geq c_2$, and thus $d(c)$ is decreasing in c .

Proof:

It is easy to see that when $\lambda_2 = 0$ the arrival process is a renewal process. Let X_c denote an arbitrary interarrival time of this renewal process. We'll prove the proposition by showing that $X_{c_1} \underset{v}{\leq} X_{c_2}$ when $c_2 \leq c_1$. To show this let M denote an exponential random variable with rate λ_1 , and let N_{c_1} denote a random variable which, conditional with M has a Poisson distribution with mean $c_1\alpha_1(M)$. Also let V_i , $i \geq 1$, denote independent (of themselves and of all the other random variables defined above) exponential random variables having rate α_2 . Then using the notation $X \overset{L}{=} Y$ to mean that X and Y have the same probability distribution, we have the following

$$(3.1) \qquad X_{c_1} \overset{L}{=} M + \frac{V_1}{c_1} + \cdots + \frac{V_{N_{c_1}}}{c_1} .$$

The above follows by interpreting E as the total amount of time the $\Lambda(t)$ process spends in its λ_1 phase during an interarrival period; N_{c_1} can be interpreted as the number of changes from λ_1 to λ_2 that occur during this interarrival period; as V_i/c_i as the time spent in the λ_2 phase during the ith visit.

Now for $c_2 \leq c_1$ we have that

$$(3.2) \qquad X_{c_2} \overset{L}{=} M + \frac{I_1 V_1}{c_2} + \cdots + \frac{I_{N_{c_1}} V_{N_{c_1}}}{c_2}$$

where the I_j are independent (of each other and of all previously defined random variables) and such that

$$I_j = \begin{cases} 1 & \text{with probability } c_2/c_1 \\[2mm] 0 & \text{with probability } 1 - \dfrac{c_2}{c_1} . \end{cases}$$

Equation (3.2) is justified by the same argument as used in (3.1) along with the fact that if each of a Poisson number of events is independently counted with probability p then the number of counted events is also Poisson.

Now it is straightforward to check that

$$\frac{I_j V_j}{c_2} \overset{>}{_v} \frac{V_j}{c_1}$$

and thus (using property 2) we see that

$$E\left[f\left(M + \frac{V_1}{c_1} + \cdots + \frac{V_{N_{c_1}}}{c_1}\right) \mid M, N_{c_1}\right]$$

$$\leq E\left[f\left(M + \frac{I_1 V_1}{c_2} + \cdots + \frac{I_{N_{c_1}} V_{N_{c_1}}}{c_2}\right) \mid M, N_{c_1}\right]$$

for all increasing convex f. It now follows from (3.1) and (3.2) by taking expectation of both sides of the above inequality, that

$$X_{c_1} \underset{v}{\leq} X_{c_2}$$

and the result follows from Proposition 1. ■

Before obtaining the limiting value of $d(c)$ let us first compute the mean and variance of the interarrival time X_c. From (3.1) we have

$$E[X_c \mid M, N_c] = M + \frac{N_c}{c\alpha_2}$$

$$Var[X_c \mid M, N_c] = \frac{N_c}{(c\alpha_2)^2}$$

and thus

$$E[X_c] = \frac{1}{\lambda_1} + \frac{\alpha_1}{\lambda_1 \alpha_2}$$

$$Var[X_c] = \frac{\alpha_1}{\lambda_1 c\alpha_2^2} + Var\left[M + \frac{N_c}{c\alpha_2}\right]$$

$$= \frac{\alpha_1}{\lambda_1 c\alpha_2^2} + E\left[\frac{\alpha_1 cM}{(c\alpha_2)^2}\right] + Var\left[M + \frac{\alpha_1 M}{\alpha_2}\right]$$

$$= \frac{\alpha_1}{\lambda_1 c\alpha_2^2} + \frac{\alpha_1}{\lambda_1 c\alpha_2^2} + \frac{1}{\lambda_1^2}\left(1 + \frac{\alpha_1}{\alpha_2}\right)^2$$

That is,

(3.3) $$E[X_c] = \frac{1}{\lambda}$$

(3.4) $$Var[X_c] = \frac{2\alpha_1}{\lambda_1 c\alpha_2^2} + \frac{1}{\lambda^2}$$

where

$$\lambda = \frac{\alpha_2\lambda_1}{\alpha_1 + \alpha_2} \; .$$

We are now ready to show that as $c \to \infty$ $d(c)$ converges to its corresponding value in the stationary case.

<u>Proposition 3</u>:

$$\lim_{c \to \infty} d(c) = \frac{\lambda E[S^2]}{2(1 - \lambda E[S])}$$

where S represents an arbitrary service time random variable.

<u>Proof</u>:

Marshall [3] showed that for GI/G/1 queues

(3.5) $$d(c) = \frac{Var[S] + Var[X_c] + (E[S] - E[X_c])^2}{2(E[X_c] - E[S])} - \frac{E[I^2]}{2E[I]}$$

where I is the length of an idle period. Now an idle period is a mixture of a random variable having the same distribution as X_c and a random variable having the same distribution as the sum of X_c and an independent exponential random variable with rate $c\alpha_2$ (the probabilities for this mixture being the respective probabilities that an idle period starts

in a λ_1 or a λ_2 phase). Hence, it follows that

(3.6)
$$X_c \underset{st}{\leq} I \underset{st}{\leq} X_c + U$$

where $\underset{st}{\leq}$ means stochastically smaller than and U is an exponential random variable with rate $c\alpha_2$ independent of the interarrival random variable X_c . Now, from the representation (3.6) and (3.3) and (3.4) we see that

$$\lim_{c \to \infty} E[I] = \frac{1}{\lambda}$$

$$\lim_{c \to \infty} E[I^2] = \lim_{c \to \infty} E\left[X_c^2\right]$$

$$= \frac{2}{\lambda^2} .$$

Letting c approach infinity in (3.5) and substituting the above yields the result. ∎

Remark:

In this section we have supposed that the $\{\Lambda(t), t \geq 0\}$ process alternates between the values λ_1 and $\lambda_2 = 0$, spending an exponential amount of time with rate $c\alpha_2$ in the $\lambda_2 = 0$ phase. However all proofs would go through in an identical fashion if we supposed that the time in the 0-phase had a distribution G_c where $G_c(x) = G(cx)$ for some arbitrary distribution G having finite variance. On the other hand, however, it is unfortunate that, to maintain renewal arrivals, the assumption of exponential times in the λ_1-phase is essential for our approach.

4. HEURISTICS AND FUTURE WORK

One of the main reasons for our belief in the conjecture of Section 1 is based on the following intuitive argument: Suppose that $[N(t), t \geq 0]$ is a nonhomogeneous Poisson Process with intensity function $\lambda(t)$, and suppose that, for some positive constant λ ,

$$\lim_{s \to \infty} \int_0^s \frac{\lambda(t)dt}{s} = \lambda \ .$$

In considering processes satisfying the above as possible arrival processes for a waiting line model we should first note that for large T they all have roughly the same distribution for the number of arrivals by time T – namely Poisson with mean roughly λT . The difference arises in the conditional distribution of the arrival times. Specifically, given that n arrivals have occurred by time T , these n arrival times have the same joint distribution as do n independent random variables chosen from the common distribution F where

$$F(x) = \begin{cases} \int_0^s \lambda(t)dt \bigg/ \int_0^T \lambda(t)dt & s \leq T \\ \\ 1 & s > T \ . \end{cases}$$

Now if we had to choose a Poisson (with mean λT) number of arrival times according to some distribution concentrated over [0,T] so as to minimize average customer waiting time then, intuitively, it seems reasonable that we would want that distribution which does not "favor"

any particular subinterval but rather has in some sense the most uniform spread. In other words we would want the uniform distribution (which has maximal entropy among all distributions concentrated on $[0,T]$). But it is precisely the uniform distribution which obtains when $\lambda(t) \equiv \lambda$.

So far due to analytical difficulties we have only been able to verify the conjecture in the special cases considered in this paper. One other possibility which may be more analytically tractable would be to consider finite capacity models. By supposing that the service distribution is exponential we can analyze such models as finite state continuous time Markov chains. In such finite capacity models the relevant quantity would no longer be the average time in queue but rather it would be the percentage of customers lost to the system due to their arrival when the system was at maximum capacity. We conjecture that this percentage of lost customers is greater than what it would be in the case of stationary arrivals.

REFERENCES

[1] Ross, S., APPLIED PROBABILITY MODELS WITH OPTIMIZATION APPLICATIONS,
       Holden-Day, (1970).

[2] Stoyan, D., "Uber einije Eijenschaften monotoner Stochastischer
       Prozesses," <u>Math. Nachr</u>., Vol. 52, pp. 21-34, (1972).

[3] Marshall, K., "Some Inequalities in Queueing," <u>Operations Research</u>,
       Vol. 16, pp. 651-668, (1968).